UKOLN

Metadata vocabularies and ontologies

Dr. Manjula Patel
Technical Research and Development
m.patel@ukoln.ac.uk

http://www.ukoln.ac.uk/

re:source The Council for Museums Archives and Libraries

Joint Information Systems Committee

UNIVERSITY OF BATH

Terminology

Metadata is

- structured data about data
- a form of language (pidgin)

A metadata vocabulary or schema:

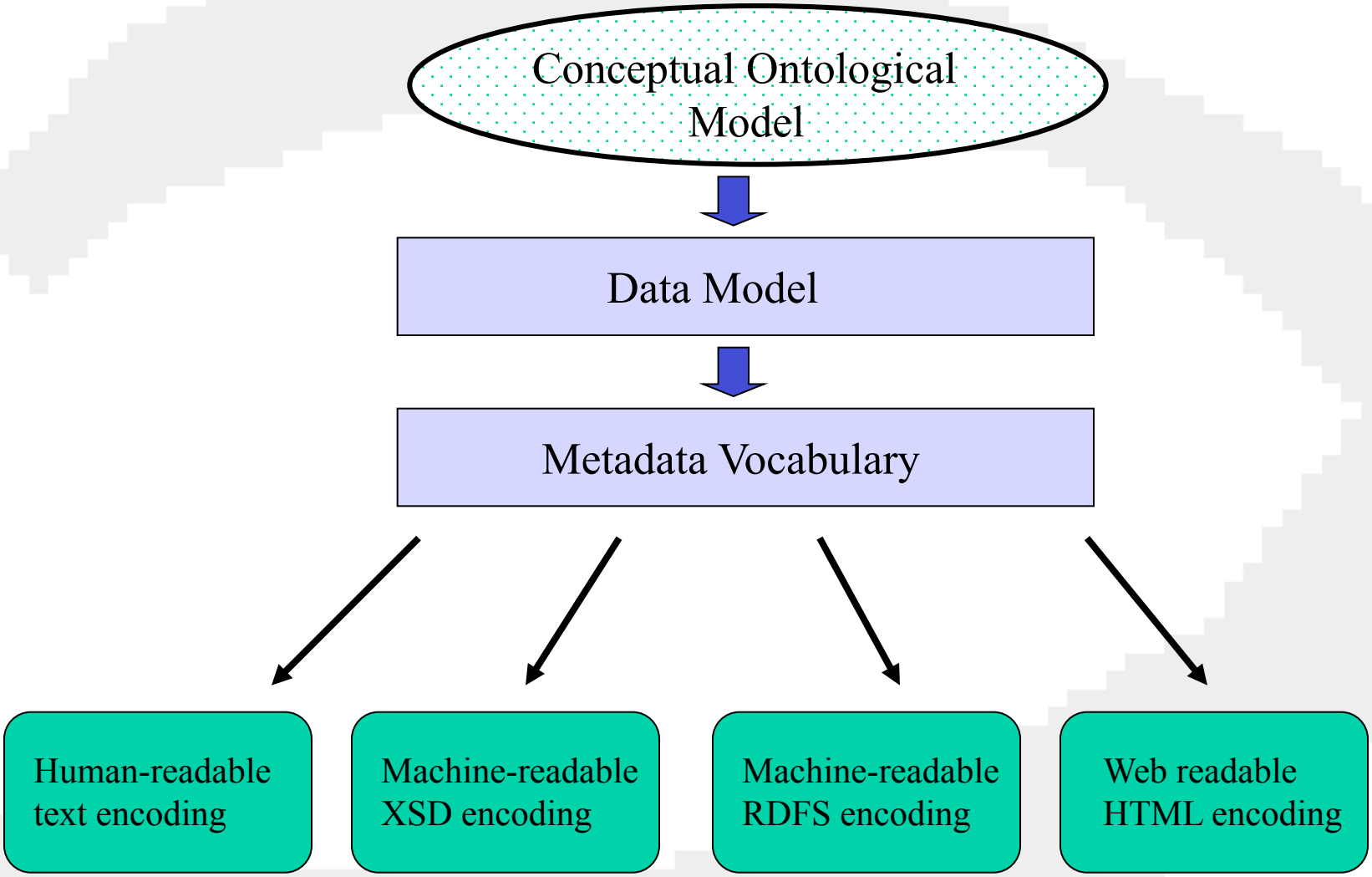
- declares a set of concepts or terms and their associated definitions and relationships
- the terms are often known as elements, attributes and qualifiers
- the definitions provide the semantics, ideally these are both human and machine readable
- in effect a manifestation of an ontology

A scheme:

- controlled vocabulary or enumerated type

UKOLN

Ontologies & Schemas



UKOLN

Types of schemas

Vocabularies range from canonical international standards to implementation specific schemas

- Single element sets
- Combinations of vocabularies
- Cross-domain
- Specific domains
- Particular applications or implementations

UKOLN

Vocabulary disclosure

Namespace schemas: declare a unique set of elements and definitions

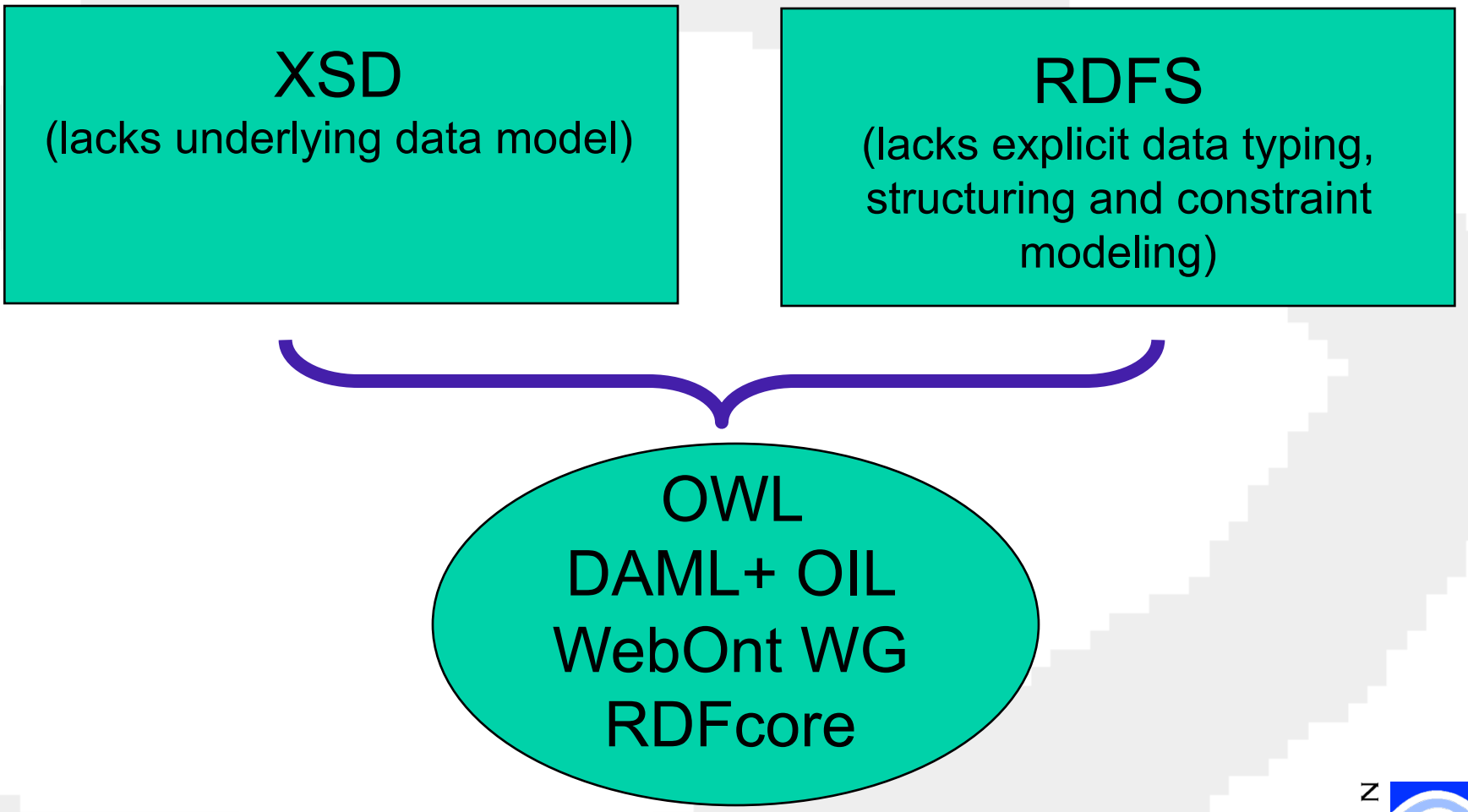
- ideally, addressed on the Web with a URI
- may be an XML or RDF schema

Application profiles: declare which terms are used by a particular application or project

- may mix-and-match terms from multiple namespaces
- may specify dependencies e.g. mandate schemes
- may adapt existing definitions for local purposes
- may declare rules for content (usage guidelines)
- may specify whether an element is mandatory, optional or repeatable

UKOLN

Encoding formats



UKOLN

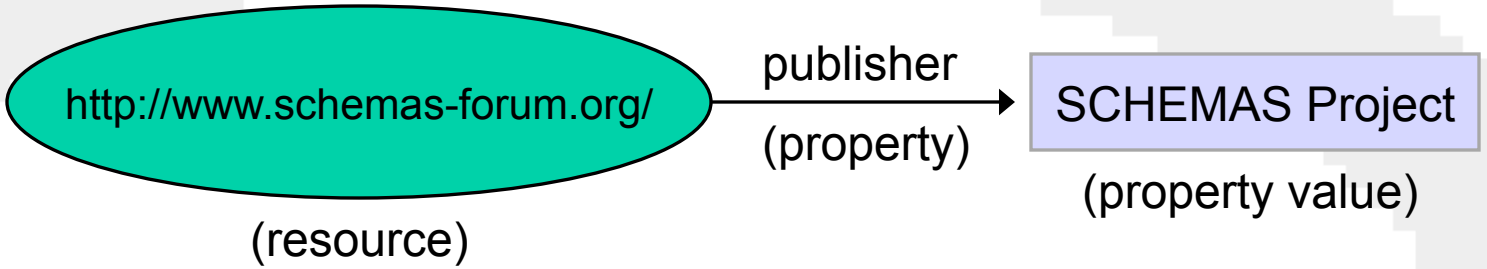
Resource Description Framework (RDF)

A shared grammar is needed to ensure that metadata is interpreted consistently

- A framework for making statements about properties and relationships of items on the Web
- A basic model for making statements about resources:
  - Resource: anything with a URI
  - Description: states the properties of the resource using terms named by URIs
  - Framework: a common model or grammar for statements
- Uses XML as serialisation syntax

UKOLN

RDF model & syntax




```
<rdf:RDF>
  <rdf:description rdf:about="http://www.schemas-forum.org/">
  <publisher>SCHEMAS Project</publisher>
  </rdf:description>
</rdf:RDF>
```

Expresses the statement:

"The SCHEMAS Project is the publisher of the resource which is identified by http://www.shemas-forum.org/"

UKOLN

RDF Schemas (RDFS)

- Web-based publication format for declaring semantics
- W3C Recommendation
- Has capability to explicitly declare semantic relations between vocabulary terms
- Machine readable, but also defines properties and classes with human readable labels and comments

Example:

Title -an element from the Dublin Core Element Set

Title: A name given to the resource (defn for humans)

dc:title (unique identifier for machine processing)

UKOLN

BIBLINK vocabulary

BIBLINK vocabulary (uses DC and BC)

DC: title, contributor, identifier, publisher, format, date, subject, description, language, rights, source

BC: creator organisation, contributor organisation, checksum, frequency, edition, place of publication, system requirements

Qualifies: title (title alternate), format (extent)

Schemes: identifier (URI, URL, DOI, ISBN, ISSN, SICI), subject (LCSH, DDC, UDC, LCC)

UKOLN

BIBLINK namespace

define a new term and associated semantics …

BIBLINK vocabulary:

Edition: A statement indicating the version or edition of the resource

```
<rdf:Property ID = "edition" >
<rdfs:label> Edition </rdfs:label>
<rdfs:comment>
  A statement indicating the version or edition of the resource
</rdfs:comment>
<rdfs:isDefinedBy
  rdf:resource =
  "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc”/>
</rdf:Property>
```

UKOLN

BIBLINK namespace

```
<rdf:Property ID="extent">
<rdfs:label> Extent </rdfs:label>
<rdfs:comment>
  The size of the resource in bytes, number of files or CD-ROMs
</rdfs:comment>
<rdfs:subPropertyOf "http://purl.org/dc/elements/1.1/format"/>
<rdfs:isDefinedBy
 "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc”/>
</rdf:Property>
```

…declare a local qualifier for dc:format

Dublin Core: Format: The format of the resource
BIBLINK : Extent: The size of the resource in bytes, no. files, or CDROMS

UKOLN

BIBLINK namespace

define a scheme for identifier …

BIBLINK vocabulary: identifier scheme: URI, URL, DOI, ISBN, ISSN, SIC

```
<rdfs:Class rdf:ID="IdentifierScheme">
  <rdfs:label> Identifier Schemes </rdfs:label>
  <rdfs:comment>
    A set of identifier encoding schemes and/or formats
  </rdfs:comment>
  <rdfs:isDefinedBy rdf:resource=
    "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc "/>
</rdfs:Class>

<rdfs:Class rdf:ID="URI">
  <rdfs:label> URI </rdfs:label>
  <rdfs:comment> Uniform Resource Locator </rdfs:comment>
  <rdfs:subClassOf rdf:resource = "#IdentifierScheme" />
  <rdfs:isDefinedBy rdf:resource =
    "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc"/>
</rdfs:Class>
```

UKOLN

BIBLINK application profile

declare reuse of terms from multiple vocabs ...

From Dublin Core:

```
<smes:uses
  rdf:resource="http://purl.org/dc/elements/1.1/description" />
<smes:uses
  rdf:resource="http://purl.org/dc/elements/1.1/language" />
<smes:uses
  rdf:resource="http://purl.org/dc/elements/1.1/rights" />
```

From BIBLINK:

```
<smes:uses rdf:resource=
 "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc#price"/>
<smes:uses rdf:resource=
 "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc#extent"/>
```

UKOLN

BIBLINK application profile

```
<smes:uses>
    <rdf:description
        rdf:about = "http://purl.org/dc/elements/1.1/title">
    <smes:comment>
        The title of the publication
    </smes:comment>
    </rdf:description>
</smes:uses>
```

…adapt the definition of dc:title

Dublin Core: Title: The title of the resource
BIBLINK : Title: The title of the publication

UKOLN

BIBLINK application profile

mandate a scheme with a specific term ...

BIBLINK vocabulary:

identifier scheme: URI, URL, DOI, ISBN, ISSN, SICI

```
<smes:uses>
  <rdf:description
      rdf:about = "http://purl.org/dc/elements/1.1/identifier">
  <rdfs:range
      rdf:resource="bc:IdentifierScheme"/>
  <rdfs:domain
      rdf:resource="http://purl.org/dc/elements/1.1/identifier"/>
  </rdf:description>
</smes:uses>
```

UKOLN

Selected references

Tim Berners-Lee, James Hendler and Ora Lassila,
The Semantic Web, Scientific American, May 2001
http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html

Rachel Heery & Manjula Patel, ***Application Profiles: Mixing and matching metadata schemas*** Ariadne, Issue 25, Sept 2000 http://www.ariadne.ac.uk/issue25/app-profiles/

Thomas Baker, Makx Dekkers, Rachel Heery, Manjula Patel, Gauri Salokhe, ***What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives***, Journal of Digital Information, October 2001
http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/

Thomas Baker, ***A Grammar for Dublin Core***
Dlib Magazine, 6(1)) October 2000

UKOLN

UKOLN

…a national focus of expertise in digital information management…

Dr. Manjula Patel
Technical Research and Development
m.patel@ukoln.ac.uk

http://www.ukoln.ac.uk/

re:source The Council for Museums Archives and Libraries

Joint Information Systems Committee

UNIVERSITY OF BATH